ANNALES
UNIVERSITATIS MARIAE CURIE-SKŁODOWSKA
LUBLIN – POLONIA
VOL. XL, 15 SECTIO H 2006

Wydział Ekonomiczny UMCS

DOROTA BARTOSIŃSKA, ANNA JANKIEWICZ-SIWEK

Metoda reprezentacyjna i jej wybrane zastosowania w badaniach społeczno-gospodarczych

Sampling and examples of some applications in social-economic surveys

Abstract: Metoda reprezentacyjna jest metodą badania częściowego, która polega na losowym wyborze próby z populacji skończonej, a wyniki z próby są uogólniane na całą populację. W życiu społeczno-gospodarczym najczęściej spotykamy takie właśnie populacje. Metoda reprezentacyjna umożliwia uzyskanie ocen parametrów zwłaszcza w takich sytuacjach, gdy populacje są bardzo liczne lub badanie ma charakter niszczący i nie jest możliwe przeprowadzenie badania pełnego. Badanie prowadzone metodą reprezentacyjną jest krótsze i tańsze w realizacji niż badanie pełne. Powodem zajęcia się metodą reprezentacyjną są jej liczne zalety, a także wiele praktycznych zastosowań. W artykule w sposób syntetyczny opisano metodę reprezentacyjną, jej rys historyczny, wybrane zastosowania w badaniach społeczno-gospodarczych w Polsce oraz błędy występujące w badaniach reprezentacyjnych.

Celem każdego badania statystycznego jest określenie wartości wybranych parametrów badanej zbiorowości. Uzyskanie poprawnych wyników badania zależy nie tylko od przyjętej metody, ale od poprawnego jej zastosowania. Pełne (wyczerpujące) badanie statystyczne polega na badaniu wszystkich jednostek tworzących populację generalną. Przeprowadzenie takiego badania nie zawsze jest możliwe, a często nawet niecelowe. Badanie pełne zbiorowości statystycznej może być kosztowne, pracochłonne, może też mieć charakter niszczący. Dlatego w praktycznych zastosowaniach dość często zachodzi konieczność ograniczenia badania do części populacji generalnej. Wśród częściowych badań statystycznych szczególne miejsce zajmuje metoda reprezentacyjna. Polega ona na losowym wyborze próby z populacji skończonej, opisie tej próby przy pomocy odpowiednich charakterystyk, a następnie uogólnieniu wyników badania na całą populację. Zastosowanie właściwego schematu losowania próby, który uwzględ-

nia możliwości techniczne, koszt i efektywność otrzymywanych wyników, pozwala uzyskać próbę reprezentacyjną. Próba reprezentacyjna jest to próba, której struktura pod względem badanej cechy nie różni się istotnie od struktury populacji generalnej.

Jeśli mówimy o metodzie reprezentacyjnej nie w ścisłym znaczeniu tego słowa, a o wnioskowaniu na podstawie zbadanej części o całości, jest ono tak dawne, jak w ogóle wiedza ludzka. W naukach przyrodniczych wszelkie wnioskowanie o zbiorowościach nieograniczenie licznych odbywało i odbywa się na podstawie zbadania niewielkiej liczby faktów. Do pierwszych przykładów takiego wnioskowania w naukach społeczno-ekonomicznych należą szacunki liczby ludności w Londynie w XVII wieku, we Francji w XVIII i w XIX wieku.[1]

Metoda reprezentacyjna oficjalnie została uznana przez Międzynarodowy Instytut Statystyczny w roku 1925. Faktyczna data jej powstania to 1985 r., gdy na sesji Międzynarodowego Instytutu Statystycznego w Bernie dyrektor Norweskiego Biura Statystycznego A. Kiaer przedstawił raport o wykorzystaniu metody badania częściowego w spisie ludności w Norwegii. Dowodził on, że badanie częściowe daje rzetelne wyniki, jeżeli próba dobrze reprezentuje populację ze względu na ważne cechy. Uważano wtedy, że reprezentatywność próby można osiągnąć przez wybór losowy albo celowy z populacji.

W roku 1913 A. L. Bowley, pierwszy profesor statystyki w London School of Economics and Political Science, użył schematu systematycznego, losując co dwudzieste gospodarstwo domowe w Reading. Obliczył również błąd średni ocenianego parametru. Przez losowy wybór próby z populacji skończonej rozumiał on wybór indywidualny z jednakowymi prawdopodobieństwami wyboru, czyli dający próbę prostą.

Znaczący wkład w rozwój metody reprezentacyjnej wniósł polski statystyk J. Spława-Neyman. Uznanie przyniósł mu artykuł przedstawiony w Królewskim Towarzystwie Statystycznym.[2] W przypadku dużej próby, a ten warunek jest zwykle spełniony w badaniach reprezentacyjnych, zalecał on losowy, a nie celowy, wybór próby. Wprowadził przedziały ufności. Sporo uwagi poświęcił losowaniu warstwowemu. Jego zdaniem nie ma potrzeby ograniczania pojęcia próby losowej do próby prostej. Losową próbę uzyskamy również, stosując wybór zespołów z populacji, a także w drodze wyboru warstwowego.

W kolejnych latach nastąpił intensywny rozwój teorii i praktyki metody reprezentacyjnej. W latach 50. ubiegłego wieku dużym zainteresowaniem cieszyły się błędy nielosowe w badaniach statystycznych. Dalszy rozwój badań nad me-

[1] S. Szulc, *Metody statystyczne*, PWE, Warszawa 1967, s. 616–617.

[2] J. Neyman, *On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection*, „Journal of the Royal Statistical Society" 1934, nr 97, s. 558–606.

todą reprezentacyjną dotyczył podejścia modelowego do skończonych struktur, złożonych populacji oraz złożonych schematów losowania.

Mimo upływającego czasu metoda reprezentacyjna wciąż cieszy się dużym zainteresowaniem. Umożliwia ona zebranie danych, niezbędnych do rozwiązywania różnorodnych problemów. Są to nie tylko problemy społeczne, przemysłowe, handlowe, ale także dotyczące innych dziedzin gospodarki, w których metoda ta nie znalazła dotychczas szerszego zastosowania. Metodę reprezentacyjną stosuje się także w badaniach eksperymentalnych, np. w rolnictwie, biologii, medycynie, technice; do statystycznej kontroli jakości w przedsiębiorstwach oraz w sondażowych badaniach opinii publicznej i konsumenckiej.

Badanie prowadzone metodą reprezentacyjną ma wiele zalet. Jest mniej pracochłonne, tańsze i krótsze w realizacji niż badanie pełne. Może dostarczyć bardziej dokładnych wyników niż badanie całkowite. Badając mniejszą zbiorowość, można więcej czasu poświęcić na staranną jej obserwację, wybrać lepszą metodę zbierania danych, dobrać bardziej odpowiedni personel do przeprowadzania badania oraz dokładniej skontrolować różne etapy badania statystycznego.

Niekiedy nie można zastąpić badania pełnego badaniem reprezentacyjnym. Z sytuacją taką mamy do czynienia, gdy zachodzi konieczność uzyskania precyzyjnych ocen parametrów dotyczących zjawisk rzadkich, gdy wyniki są opracowywane dla małych części zbiorowości generalnej lub gdy tablice przedstawiające wyniki badania muszą być szczegółowe.[3]

Metoda reprezentacyjna posiada wiele różnorodnych zastosowań. Może zastępować spisy pełne. W spisach próbnych służy do sprawdzenia projektów formularzy i instrukcji spisowych. Wykorzystywana jest do rozszerzania tematyki spisu pełnego. Pomaga także szybko uzyskać wstępne wyniki tego spisu. Stosowana jest do badań kontrolnych błędów nielosowych, a także do badań o charakterze niszczącym.

Obecnie metoda reprezentacyjna jest szeroko stosowana w statystyce publicznej na całym świecie. Badania prowadzone tą metodą obejmują różne dziedziny życia społeczno-gospodarczego. Na podstawie najbardziej aktualnych publikacji Głównego Urzędu Statystycznego przedstawiono w tabeli 1 te badania społeczno-gospodarcze prowadzone w Polsce, których metodologia jest dostępna dla odbiorców danych statystycznych. Większość tych badań przeprowadzanych jest z częstotliwością roczną. Sezonowe zmiany występujące w wielu dziedzinach życia społeczno-gospodarczego wpływają na wzrost częstotliwości badań, np. badanie aktywności ekonomicznej ludności przeprowadzane jest co kwartał, badanie pogłowia trzody – trzy razy w roku, a badanie pogłowia bydła, owiec i drobiu – co pół roku.

W rozpatrywanych badaniach reprezentacyjnych badane populacje obejmują ogół przedsiębiorstw, gospodarstw domowych, pracowników, ludności lub in-

[3] R. Zasępa, *Metoda reprezentacyjna*, PWE, Warszawa 1972, s. 21.

dywidualnych gospodarstw rolnych w Polsce. Operaty losowania prób do badań reprezentacyjnych są tworzone na podstawie spisów pełnych i różnych rejestrów. W niektórych badaniach jednostki badania nie pokrywają się z jednostkami losowania. Zbiór jednostek badania tworzy populację generalną. Jednostki te mają pewne własności wspólne, które pozwalają rozstrzygnąć, czy określona jednostka należy do danej populacji czy też nie; różnią się też tymi cechami, które są przedmiotem badania. Natomiast operat losowania próby to wykaz jednostek losowania, który powinien spełniać następujące warunki[4]:

- kompletności, czyli powinien obejmować wszystkie jednostki populacji;
- aktualności, czyli informacje zawarte w operacie powinny być aktualne;
- identyfikowalności, czyli musi istnieć możliwość odszukania wylosowanej jednostki.

W większości badań w celu uzyskania próby reprezentatywnej przeprowadzane jest warstwowanie przed losowaniem. Losowanie warstwowe polega na tym, że przed przystąpieniem do losowania są tworzone warstwy, w taki sposób, by każdy element populacji był zaliczony do jednej i tylko jednej warstwy i by nie było elementów pozostających poza warstwami. Następnie losowania próby dokonuje się z każdej warstwy w taki sposób, by losowania z różnych warstw były wzajemnie niezależne. Wylosowane ze wszystkich warstw elementy składają na ostateczną próbę losową.[5]

W przypadku losowania prostego bez zwracania każda z prób ma takie same szanse, aby zostać wylosowaną. Może się zdarzyć, że otrzymana próba będzie niereprezentatywna, tzn. taka, w której pewne jednostki nie występują, mimo że udział takich jednostek w populacji jest znaczny.[6] Warstwowanie umożliwia wyeliminowanie tych prób możliwych do wylosowania w przypadku losowania nieograniczonego, które są mało reprezentatywne. Do próby mogłyby wejść na przykład tylko jednostki małe albo tylko jednostki duże. Próby takie byłyby niereprezentatywne.

W rozpatrywanych badaniach są stosowane dwa schematy losowania warstwowego – proporcjonalny i optymalny. Optymalnym schematem losowania warstwowego jest schemat, w którym przy danym całkowitym koszcie badania wariancja estymatora jest najmniejsza. Natomiast schemat losowania warstwowego proporcjonalnego polega na tym, że liczebności próby w warstwie są proporcjonalne do populacyjnych liczebności warstw. W dwóch badaniach: aktywności ekonomicznej ludności oraz budżetów gospodarstw domowych jest stosowany schemat losowania dwustopniowego. Polega on na tym, że najpierw losuje się

[4] C. Bracha, *Teoretyczne podstawy metody reprezentacyjnej*, Wydawnictwo Naukowe PWN, Warszawa 1996, s. 26–27.

[5] Z. Pawłowski, *Wstęp do statystycznej metody reprezentacyjnej*, PWN, Warszawa 1972, s. 74.

[6] C. Bracha, *op. cit.*, s. 102.

według wybranego schematu pewną liczbę jednostek (jest to losowanie pierwszego stopnia), a następnie w wylosowanych jednostkach dokonuje się losowania jednostek zgodnie z określonym schematem (jest to losowanie drugiego stopnia).[7]

Po wybraniu schematu oraz operatu losowania jest pobierana próba o określonej liczebności (patrz tab. 1). Wielkość próby zależy między innymi od nakładów przeznaczonych na badanie statystyczne. Na podstawie wyników uzyskanych z próby szacuje się parametry populacji generalnej. W przypadku opisywanych tu badań reprezentacyjnych szacowanymi parametrami są: frakcja (np. wskaźnik zatrudnienia, tj. udział pracujących w liczbie ludności ogółem), średnia (np. przeciętna płaca brutto) i wartość globalna (suma) badanej cechy (np. liczba bezrobotnych, powierzchnia zasiewów) oraz iloraz dwóch wartości globalnych (np. plony). Oceny tych parametrów są obliczane dla podstawowych przekrojów kraju jako całości oraz dla poszczególnych województw. Do szacowania parametrów są stosowane estymatory bezpośrednie warstwowe (w przypadku losowania warstwowego jednostopniowego) lub złożone „wielopiętrowe" estymatory zaliczające się do estymatorów typu ilorazowego (w przypadku losowania wielostopniowego). W obu przypadkach wyniki uzyskane z próby przeważa się mnożnikami uogólniającymi, które są równe odwrotnościom prawdopodobieństw wyboru danej jednostki do próby. Mnożnik uogólniający informuje o tym, ile jednostek populacji reprezentuje dana jednostka wylosowana do próby.

Liczebność próby, przyjęty schemat losowania oraz zastosowane metody estymacji wpływają na wielkość błędów losowych w badaniu reprezentacyjnym. Błędy losowe występują tylko w badaniach reprezentacyjnych i są związane z losowaniem próby. Powstają stąd, że dane uzyskuje się tylko od jednostek wylosowanych do próby, a nie od wszystkich jednostek populacji. Miarą błędu losowego jest wariancja estymatora, wyrażona jako kwadrat różnicy między oceną parametru z próby a jego wartością oczekiwaną ze wszystkich możliwych prób. Gdyby wielokrotnie losowano próby o takiej samej liczebności, w takich samych warunkach badania i obliczano na ich podstawie oceny parametru, to oceny te ułożyłyby się dookoła ich wartości oczekiwanej zgodnie z rozkładem normalnym. W praktyce jednak losowana jest tylko jedna próba. Jesteśmy w stanie z tej jednej próby ocenić błąd losowy, gdyż zmienność w próbie między jednostkami odzwierciedla zmienność między różnymi próbami. W publikacjach z wynikami badań reprezentacyjnych miarą błędu losowego są współczynniki zmienności estymatora podawane w procentach dla kilku wybranych cech w kraju. Ich wielkość zależy od zmienności badanej cechy, z reguły nie przekraczają one 3% dla kraju, chociaż zdarzają się też większe wartości.

Oprócz błędów losowych, w każdym badaniu zarówno reprezentacyjnym, jak i pełnym występują błędy nielosowe. Błędy te nie są wynikiem losowania.

---

[7] *Ibid.*, s. 136.

Powstają w poszczególnych etapach badania statystycznego. Błędy nielosowe to błędy w kompletności i błędy w treści. Błędy w kompletności powstają w wyniku popełnianych pomyłek w zakresie opuszczenia, powtórzenia, błędnego włączenia jednostek populacji (są to błędy pokrycia) oraz w wyniku nieuzyskania danych (niezlokalizowanie, nieobecność w domu, brak kontaktu, odmowy, zagubione kwestionariusze, kwestionariusze odrzucone w czasie kontroli). Natomiast błędy treści to: błędy odpowiedzi (błąd respondenta, błąd ankietera, błędy zapisu i kopiowania); błędy opracowania danych (błędy kodowania, wprowadzania danych, redagowania, tabulacji i obliczeń); błędy analizy i prezentacji wyników (błędne metody analizy, błędy interpretacji, błędne wnioskowanie, błędna prezentacja wyników).[8] Wielkość błędów nielosowych jest zazwyczaj dużo wyższa od błędów losowych. Przykładowo w 2004 r. wskaźnik braku odpowiedzi w badaniu kosztów pracy wyniósł 20%, w tym odmowy aż 13%. W badaniu budżetów gospodarstw domowych wskaźnik braku odpowiedzi był większy od 40%. W badaniu najmniejszych przedsiębiorstw błędy w kompletności osiągnęły wielkość 41,6%.

W celu eliminacji błędów nielosowych stosowane są różne metody. Na uwagę zasługują metody bezpośrednie, które stosowane są na różnych etapach badania statystycznego. Do nich zaliczamy badania kontrolne w czasie prowadzenia badania lub bezpośrednio po jego przeprowadzeniu. Druga grupa metod to metody pośrednie, które stosowane są wtedy, gdy wyniki badania zostaną już opracowane, a nawet opublikowane. Nie wymagają żadnej dodatkowej pracy terenowej, a jedynie wykorzystują już istniejące i dostępne dane z różnych źródeł.

Rozmiary błędów losowych i nielosowych decydują o dokładności badania reprezentacyjnego. Dokładność informacji statystycznej wyraża jej bliskość z wartością prawdziwą. Wartość prawdziwa, która w praktyce prawie nigdy nie występuje, jest wartością, którą otrzymano by, gdyby dane były zebrane i opracowane bez żadnych błędów dla wszystkich jednostek populacji. Wynik badania byłby dokładny, gdyby w czasie przeprowadzenia pełnego badania statystycznego były idealnie spełnione wszystkie istotne warunki badania, takie jak: przedmiot badania, sytuacja społeczno-polityczna kraju, metody zbierania danych, zastosowane procedury statystyczne, w tym metody estymacji, kwalifikacje personelu biorącego udział w badaniu (ankieterów, kontrolerów, osób kodujących, wprowadzających dane itp.), okres badania, sposób ujęcia pytań i wyjaśnień, definicje pojęć, sposób opracowania zabranego materiał itp. Dokładność statystyczna jest najważniejszą cechą jakości danych statystycznych, rozumianej jako użyteczność wyników badań statystycznych.

Podsumowanie:

1. Metoda reprezentacyjna może być stosowana w przypadku bardzo licznych populacji generalnych, gdy badanie całkowite jest niemożliwe, kosztowne bądź ma charakter niszczący.

[8] J. Kordos, *Dokładność danych w badaniach społecznych*, GUS, Warszawa 1987, s. 24–25.

2. Badanie prowadzone metodą reprezentacyjną może dostarczyć znacznie dokładniejszych danych niż badanie pełne.
3. Zakres zastosowań metody reprezentacyjnej jest bardzo szeroki. Są to badania społeczne, rolnicze, gospodarcze i inne.
4. W badaniach prowadzonych metodą reprezentacyjną występują błędy losowe i nielosowe. Należy dążyć do ich minimalizacji, bo od ich wielkości zależy użyteczność wyników badania statystycznego.
5. Zastąpienie badań całkowitych przez badania reprezentacyjne nie zawsze jest możliwe do wykonania.

Tab. 1. Opis metodologii wybranych badań prowadzonych metodą reprezentacyjną w Polsce
Description of methodology of some sample surveys conducted in Poland

Lp.	Nazwa badania	Częstotliwość badania	Badana populacja	Liczebność próby (% populacji)	Schemat losowania próby	Operat losowania
No.	Name of survey	Frequency of survey	Tested population	Sample size (in per cent of population)	Sampling selection scheme	Sampling frame
1	Badanie aktywności ekonomicznej ludności	Co kwartał	Ludność w wieku 15 lat i więcej	45,9 tys. osób (1,5%)	Dwustopniowy (1 st. – rejony statystyczne w miastach i obwody spisowe na wsiach, 2 st. – mieszkania)	1 st. – Krajowy Rejestr Podziału Terytorialnego Kraju, 2 st. – wykazy mieszkań
2	Badanie kosztów pracy	Co 4 lata	Jednostki gospodarki narodowej zatrudniające 10 osób lub więcej	28 tys. jednostek (20%)	Warstwowy proporcjonalny	Wykaz jednostek gospodarki narodowej zatrudniających 10 osób lub więcej
3	Badanie wynagrodzeń według zawodów	Co rok	Pracownicy pełnozatrudnieni	0,6 mln pracowników (9,7%)	Zespołowy	Wykaz jednostek gospodarki narodowej zatrudniających 10 osób lub więcej
4	Badanie budżetów gospodarstw domowych	Co rok	Gospodarstwa domowe	32,2 tys. gospodarstw domowych (0,3%)	Dwustopniowy (1 st. – terenowe punkty badań, 2 st. – mieszkania)	1 st. – wykazy rejonów statystycznych opracowane do NSP, 2 st. – wykazy mieszkań

5	Badanie najmniejszych przedsiębiorstw	Co rok	Przedsiębiorstwa zatrudniające do 9 osób	110,9 tys. przeds. (3,6%)	Warstwowy	Utworzono na podstawie Bazy Jednostek Statystycznych i zaktualizowano informacjami z Ministerstwa Finansów
6	Badanie warunków powstania i działania oraz perspektyw rozwojowych polskich przedsiębiorstw	Co rok	Przedsiębiorstwa nowo utworzone zatrudniające do 9 osób	2 tys. przeds. (1,1%)	Warstwowy proporcjonalny	Utworzono na podstawie REGON
7	Badanie użytkowania gruntów, powierzchni zasiewów, pogłowia zwierząt i charakterystyki gospodarstwa rolnego	Co rok	Indywidualne gospodarstwa rolne	200 gosp. (6,9%)	Warstwowy optymalny	Utworzono na podstawie wyników NSP i PSR
8	Badanie pogłowia bydła, owiec i drobiu	2 razy w roku	Indywidualne gospodarstwa rolne posiadające bydło, drób lub owce oraz o powierzchni użytków rolnych 15 ha i więcej	20 tys. gosp. (1,5%)	Warstwowy optymalny	Utworzono na podstawie wyników NSP i PSR
9	Badanie pogłowia trzody chlewnej	3 razy w roku	Indywidualne gospodarstwa rolne posiadające trzodę oraz o powierzchni użytków rolnych 15 ha i więcej	15 tys. gosp. (1,8%)	Warstwowy optymalny	Utworzono na podstawie wyników NSP i PSR
10	Badanie plonów zbóż	Co rok	Indywidualne gospodarstwa rolne	11,6 tys. gosp. (0,4%)	Warstwowy optymalny	Utworzono na podstawie wyników NSP i PSR

Źródło: Opracowanie własne na podstawie publikacji Głównego Urzędu Statystycznego.

SUMMARY

Sample surveys consist on probability selection of samples from finite populations and results from samples are generalized on the whole population. The most frequently met populations in our social-economic reality are the finite ones. Sample surveys allow to obtain estimates of parameters especially when a population is very large or when as a result of testing all the considered units are destroyed and it is impossible to conduct a full-scale survey. Sample survey takes less time and is cheaper than full-scale survey. We have taken up the sampling because sample surveys have a lot of advantages and many practical applications. In this paper we presented methodology of sample surveys, short history of sampling and some of its applications in social-economic surveys. We dealt with populations, sample sizes, sampling selection schemes, sampling frames and methods of estimation in social-economic surveys conducted in our country. We presented both sampling and non-sampling errors occurring in sample surveys.